Salient in the mind, salient in prosody

Kaland, C.C.L.; Krahmer, E.J.; Swerts, M.G.J.

Published in:
Proceedings of the 33rd Annual Conference of the Cognitive Science Society

Publication date:
2011

Document Version
Publisher's PDF, also known as Version of record

Salient in the mind, salient in prosody

Constantijn Kaland (c.c.l.kaland@uvt.nl)
Emiel Krahmer (e.j.krahmer@uvt.nl)
Marc Swerts (m.g.j.swerts@uvt.nl)
Tilburg centre for Cognition and Communication (TiCC), The Netherlands

Abstract

Research in psychology has shown that when people are told not to think about a pink elephant they cannot avoid doing just that. Similar results are found for language production in that people leak information about hidden figures when instructed to ignore those figures. It is argued that the salience of information plays a crucial role in these effects. The present study investigates how different factors of salience affect speakers' lexical and prosodic behaviour. Results indicate that those factors affect lexical use and prosody in different ways and, crucially, that adjectives signalling leaked information are prosodically more prominent, as measured by both by acoustic analysis and prominence ratings.

Keywords: prosody; ironic processes; speech production.

Introduction

A central problem in recent research on speech production relates to the question to what extent speakers take into account what their listeners know or do not know. Quite a few studies report that speakers, under certain conditions, tend to violate the Gricean maxim of quantity (Grice, 1975) by giving more information than is strictly speaking necessary from the perspective of a listener. Such evidence is often provided by studies that use referential communication tasks where some information is shared by all interlocutors (common ground; Clark, 1996) and some information is only available for the speaker (privileged ground).

From the use of adjectives there is evidence that speakers sometimes fail to adapt to listeners when describing pictures (Wardlow Lane, Groisman, & Ferreira, 2006). For example, to describe a mutually visible figure speakers may say "the small triangle", even when a big triangle is occluded for the listener (as in Figure 1). Interestingly, when speakers are instructed not to give information about the occluded figure, the target figure is even more often described with an adjective that refers to the contrast between the target and the occluded figure. Wardlow Lane et al. (2006) concluded that speakers leak privileged information about their own perspective.

These findings are explained according to the theory of ironic processes by Wegner (1994). This theory assumes two cognitive processes; an operator process which is responsible for running actions and a monitor process which is in constant search for failures of the first. Instructions of the type "do not..." cause the speaker to be aware of unsuccessful scenarios, which then triggers counter-behaviour. Thus, ironic processes seem to boost the salience of a contrast relation between two figures so that a speaker is more likely to use an adjective referring to that contrast. The current paper builds on previous work to tackle a number of issues that remain unexplored. In particular, we explore questions related to the salience of the information that is described, with respect to both the lexical and prosodic characteristics of adjectives.

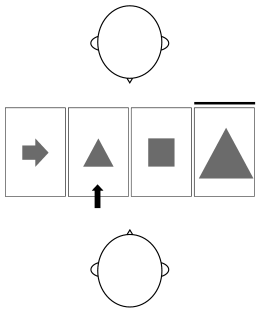

Figure 1: Experimental setup of Wardlow Lane et al. (2006). The arrow indicates the target object, the bar indicates the occluder.

First, it remains to be seen whether the effects found by Wardlow Lane et al. (2006) generalize to all kinds of adjectives. As is known from theories of incremental speech production, a speaker may start articulating without being aware of certain contrast relations (Fry, 1969; Levelt, 1989; Pechmann, 1989). This claim accounts for the general finding in object naming that size adjectives are used less often than color adjectives. That is, size refers to a feature relative to another object, whereas color can be named when only one object is known for the speaker. In Wardlow Lane et al. (2006) only size was investigated. One could question whether prohibitive instructions have an effect on naming features that are already highly salient, such as color.

Second, it may also matter what kind of contrast relation exists between different objects. Indeed, the information an adjective gives about a contrast can vary. That is, *big* in "the big triangle" is more informative with respect to a privileged smaller triangle than with respect to a privileged smaller square, the latter differing in two features from the target. The setup of Wardlow Lane et al. (2006) does not allow for such variation: the speaker either leaks information (i.e. by naming the adjective) or not. We hypothesize that contrasts consisting of one feature are more salient than contrasts consisting of two features, since the former have fewer similarities between the objects. So, it can be expected that the more salient a contrast is, the more likely speakers are to refer to it with an adjective.

And third, by looking at the adjectives' frequency of occurrence, Wardlow Lane et al. (2006) claim that ironic processes affect the grammatical encoding stage in speech production (Levelt, 1989) such that speakers leak information in their utterance. However, Pechmann (1984, 1989) shows that speakers sometimes redundantly use adjectives (overspecification). Therefore, just by looking at its occurrence we cannot tell whether an adjective implicitly refers (i.e. leaks) or is used redundantly. That is, leaking implies that speakers encode the hidden information in their utterance such that listeners can pick it up. By investigating the prosodic realisation of adjectives we can shed light on this question. It is known from Pechmann (1984) that under normal circumstances (i.e. when there is no privileged information) the adjective signalling a contrast within a certain visual context, as in Figure 1, is not likely to be prosodically marked. Incremental production strategies account for that; speakers may start articulating before they have a full cognitive representation of a visual context. Only when such a representation is available, i.e. with respect to a previous contrastive context, the adjective is always marked prosodically by means of a pitch accent. The experimental setup of Pechmann (1984), however, did not include a setting in which the salience of one hidden member of the contrast was boosted by ironic processes. It remains to be seen whether ironic processes are strong enough to affect both the formulation *and* articulation of an utterance, which is expected if one assumes that speakers indeed leak information.

To shed light on the issues mentioned above, the current study investigates how factors of salience, such as the communicative setting, the type of contrast and the contrastive feature affect both the lexical use and prosodic realisation of adjectives. A production task elicits speakers' utterances which are analysed in terms of frequency of occurrence, and acoustically and perceptually to explore prominence patterns.

Recording procedure

Utterances are collected following the paradigm of Wardlow Lane et al. (2006) where speakers have to describe figures for their listeners (Figure 1). In the present study the likelihood of uttering an adjective to refer to a contrast is manipulated by three factors related to the salience of a contrast: communicative setting (shared, privileged, conceal), contrast (single, double) and feature (size, color). First, as for *communicative setting* we replicate the experimental conditions of Wardlow Lane et al. (2006) and included an additional condition. As for the added condition, the figures forming a contrast are accessible for both the speaker and the listener such that the adjective had to be named to prevent ambiguity (shared setting). The shared setting most closely resembles the experimental setup of Pechmann (1984) in that there is no privileged information. To replicate the effect Wardlow Lane et al. (2006) find, their baseline condition (henceforth privileged setting) and conceal condition (henceforth conceal setting) are included. In those settings one of the figures in the stimulus is occluded for the listener. The privileged and conceal setting differ in that speakers are instructed not to give information about the occluded figure in the conceal setting. In the privileged setting no such instruction is given. Second, the type of *contrast* differs such that two figures in a stimulus are distinguishable by either one feature or two features. For example, the target figure can be a small triangle contrasting with a big triangle (single contrast) or the target figure can be a small diamond contrasting with a big flash (double contrast), see Figure 2. The latter contrast is assumed to be less salient than the former. Third, the contrastive *feature* is either size or color. For example, the contrast concerns a small and big triangle (size) or a red and a grey cross (color). In the double contrasts the second contrastive feature is always shape. That is, for example, the contrast concerns a small diamond and a big flash (size) or a green star and a black circle (color), see Figure 2. Four instances of each stimulus are created such that different shapes and colors are used. Note that for all size contrasts figures have the same color whereas for all color contrasts figures have the same size (Figure 2). In total, the experiment consists of 48 stimuli (3 communicative settings, 2 contrasts, 2 features, 4 repetitions).

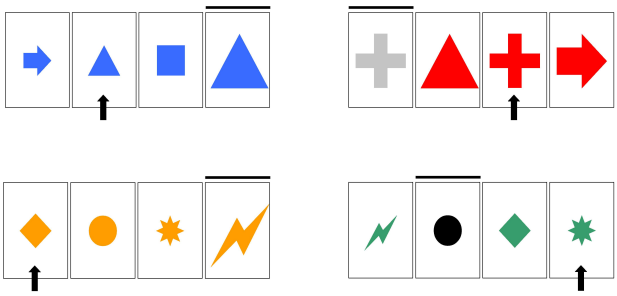

Figure 2: Example stimuli per contrast (top: single, bottom: double) and feature (left: size, right: color). The arrow indicates the target object, the bar indicates the occluder.

Each stimulus is prevented from contrasting minimally with the previous stimulus, which would affect prosodic marking (Pechmann, 1984). Therefore, two successive targets are chosen from two different sets of shapes and colors. That is, if one target is a blue circle the following target is never blue colored or circle shaped (e.g. a red arrow).

A total of 42 different participants acted as speaker (31 women, 11 men, M_{age} = 21.3 years, age range: 18-29 years). The same number of participants acted as listener. All participants were native speakers of Dutch and students at Tilburg University who had to take part as a course requirement.

Table 1: Effects of communicative setting, contrast and their interactions on repeated proportion measures of size and color adjective use after analysis of variance (matching: upper left / lower right; mismatching: upper right / lower left).

Feature	Factor	Size adjectives	Color adjectives
Size	Setting	$F(2,82) = 22.74$, $p < .001$, $\eta_p^2 = .36$	$F(2,82) = 1.21$, n.s.
	Contrast	$F(1,41) = 74.45$, $p < .001$, $\eta_p^2 = .64$	$F(1,41) < 1$, n.s.
	Setting*Contrast	$F(2,82) = 32.48$, $p < .001$, $\eta_p^2 = .44$	$F(2,82) < 1$, n.s.
Color	Setting	$F(2,82) = 2.65$, n.s.	$F(2,82) = 12.48$, $p < .001$, $\eta_p^2 = .23$
	Contrast	$F(1,41) = 1.13$, n.s.	$F(1,41) = 28.00$, $p < .001$, $\eta_p^2 = .41$
	Setting*Contrast	$F(2,82) < 1$, n.s.	$F(2,82) = 22.27$, $p < .001$, $\eta_p^2 = .35$

Adjective frequency analysis

Method

Adjective use is calculated separately for size and color as a proportion so that 1 means that all participants uttered an adjective in all four instances of a stimulus. A distinction is made between adjectives that match and mismatch the contrastive feature in the stimulus. That is, whenever speakers use a size adjective when the contrastive feature is size (or a color adjective when the feature is color) adjectives are called *matching*. Whenever speakers use a size adjective when the contrastive feature is color (or vice versa) adjectives are called *mismatching*. Utterances including both adjectives count once for size and once for color (one matching and one mismatching). Four analyses of Variance (ANOVAs) are performed with repeated mean proportion measures of adjective use as dependent variables (i.e. size and color, matching and mismatching) and with communicative setting (3 levels: shared, privileged, conceal) and contrast (2 levels: single, double) as within subject factors.

Results

A general effect of adjective type is found in that color adjectives (M = .61) are uttered more often than size adjectives (M = .42): [$t(2015) = 15.76$, $p < .001$]. Main effects of contrast and communicative setting are significant only for adjectives that match the contrastive feature (cf. Table 1 and Figure 3). That is, speakers are more likely to utter a matching adjective referring to a single contrast (M_{size} = .66, M_{color} = .75) than to a double contrast (M_{size} = .37, M_{color} = .59). As for setting, pairwise comparisons reveal that speakers use fewer matching adjectives in the privileged setting (M_{size} = .36, M_{color} = .56), both compared to the shared setting (size: M = .65, $p < .001$; color: M = .78, $p < .001$) and to the conceal setting (size: M = .52, $p < .01$; color: M = .69, $p < .05$). The interaction effect between setting and contrast reveals that the difference between the types of contrast is significantly larger in the shared than in the privileged or conceal setting. This effect can be related to the disambiguating function the matching adjective has in the shared setting for single contrasts. In this situation uttering only a noun would underspecify the target. Speakers are mostly aware of this fact, as shown by proportion values reaching 1 (Figure 3, top).

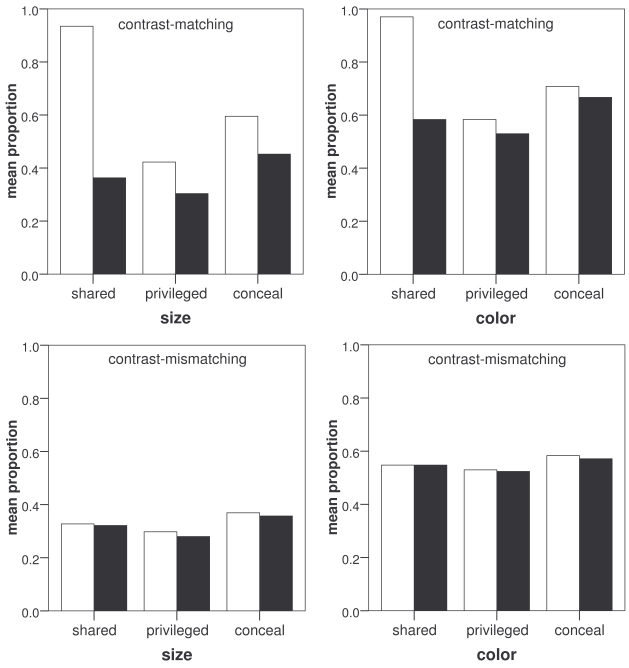

Figure 3: Mean proportions contrast-matching (top) and contrast-mismatching (bottom) size (left) and color (right) adjectives as a function of communicative setting and contrast (white = single, black = double).

Speakers use more matching adjectives in the conceal setting (compared to privileged), which confirms the findings of Wardlow Lane et al. (2006). To zoom in on this effect, ANOVAs are performed for each level of the factors contrast and feature on repeated proportion measures of size and color adjective use as dependent variables and with communicative setting as within subject factor. Pairwise comparisons reveal a significant increase between the privileged and conceal setting both for matching size and matching color adjectives either referring to a single contrast (color marginally) or to a double contrast (Table 2).

The main effect sizes are larger for size than for color adjectives (Table 1), indicating that a boost in salience affects low salient features more than high salient features. The overall high rate of adjectives could be the result of a size or color contrast in all stimuli, whereas in Wardlow Lane et al. (2006) there is a size contrast in their test stimuli (M = .1), but not in their control stimuli (M = .009).

Table 2: Mean differences, standard errors and 95% confidence intervals (after pairwise comparisons: conceal–privileged) for proportion measures of size and color adjective use per contrast (white = single, shaded = double) and feature, with [a] = $p < .001$, [b] = $p < .01$, [c] = $p < .05$ and [d] = n.s. after Bonferroni correction.

Feature	Size adjectives		Color adjectives	
	MD (*SE*)	95% CI	*MD* (*SE*)	95% CI
Size	.17[a] (.05)	(.08, .26)	.05[d] (.05)	(-.05, .16)
	.15[b] (.05)	(.05, .25)	.05[d] (.05)	(-.06, .15)
Color	.07[d] (.05)	(-.03, .17)	.13[b] (.04)	(.04, .21)
	.08[d] (.05)	(-.02, .18)	.14[c] (.05)	(.03, .24)

In sum, the present results reveal that communicative setting, contrast and feature contribute additively to the likelihood that an adjective is uttered to refer to that contrast. The different variables only affect the use of contrast-matching adjectives. A fair amount of contrast-mismatching adjectives confirms that speakers have a tendency to use adjectives as overspecification (Pechmann, 1989). The next section will explore how the salience factors influence the prosodic realisation of adjectives.

Acoustic analysis

Acoustic analysis concerns pitch, which is believed to be the most important correlate of accent (Collier & 't Hart, 1981) and duration, which correlates strongly with redundancy (Lieberman, 1963; Aylett & Turk, 2004). Utterances consist of contrast-matching adjectives taken from the shared and the conceal setting, as only those adjectives' occurrence is affected by the factors communicative setting and contrast.

Method

Analysis is done on utterances including one adjective and a noun; 104 utterances for size, either *kleine* (small) or *grote* (big), and 121 for color, either *groene* (green) or *grijze* (gray). Utterances including monosyllabic adjectives as a result of Dutch inflection rules are excluded from analysis. Furthermore, utterances do not include fillers such as "uhhmm". An additional 4 utterances in which vowel reduction does not allow for F0-measurement are excluded from pitch analysis (3 for size and 1 for color). Utterances are segmented manually by auditory perception and spectral analysis in Praat (Boersma & Weenink, 2010). Using a script, the maximum F0 (Hz) and segment durations are extracted. To abstract over gender differences and to better represent perceived prominence by the human ear Hertz values are converted into ERB values using the formula by Glasberg & Moore (1990) where f is the value in Hertz: $[21.4*\log_{10}(0.00437*f+1)]$. The pitch value of the noun is subtracted from the pitch value of the adjective resulting in a relative measure which accounts for the fact that accents are perceived relative to each other. Durations are computed relative to the whole utterance, which abstracts over speech tempo differences among participants and over different noun lengths.

Univariate ANOVAs are performed with relative pitch and relative duration measures of size and color adjectives (both separately and taken together) as dependent variables and with communicative setting (2 levels: shared, conceal) and contrast (2 levels: single, double) as within-subject factors.

Results

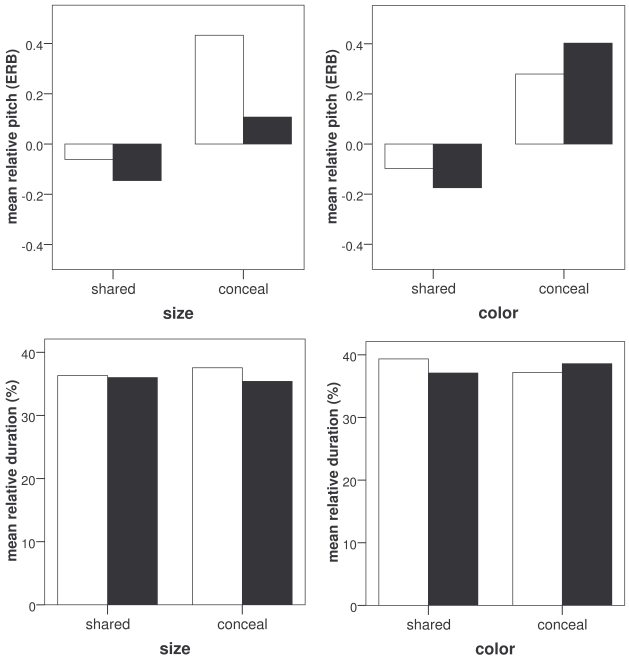

Figure 4: Mean relative pitch (top) and mean relative duration (bottom) for size (left) and color (right) adjectives as a function of communicative setting and contrast (white = single, black = double).

No overall differences are found between size and color adjectives; neither for relative duration (M_{size} = 36.45, M_{color} = 38.34, $t(223) = -1.67$, $p = .09$) nor for relative pitch (M_{size} = .04, M_{color} = .08, $t(219) = -.33$, $p = .75$). Results show that adjectives in the conceal setting are uttered with a higher relative pitch (M_{size} = .35, M_{color} = .38) than in the shared setting (M_{size} = -.08, M_{color} = -.14), see Table 3 and Figure 4. No effects are found for contrast. The effect of communicative setting remains significant after analysis on the data of both size and color adjectives: [$F(1,217) = 16.21$, $p < .001$, $\eta_p^2 = .06$]. None of the analyses reveal significant effects on relative duration (Table 3, Figure 4).

Although speakers produce adjectives in the conceal setting with a higher relative pitch, it is unclear whether this effect is strong enough to contribute to listeners' perception of prominence. This question will be addressed in a rating task.

Table 3: Effects of communicative setting and contrast on measures of relative pitch and relative duration of size and color adjectives after univariate analyses of variance.

Measure	Factor	Size adjectives	Color adjectives
Relative pitch	Setting	$F(1,97) = 5.09$, $p < .05$, $\eta_p^2 = .05$	$F(1,116) = 8.51$, $p < .01$, $\eta_p^2 = .06$
	Contrast	$F(1,97) = 1.42$, n.s.	$F(1,116) < 1$, n.s.
Relative duration	Setting	$F(1,100) < 1$, n.s.	$F(1,117) < 1$, n.s.
	Contrast	$F(1,100) < 1$, n.s.	$F(1,117) < 1$, n.s.

Prominence judgments

Utterances used for acoustic analysis (n = 225) are presented to listeners in a prominence rating task. A total of 13 participants (10 men, 3 women, M_{age} = 29.8 years, age range: 24-44) completed the task. All of them were native speakers of Dutch without hearing problems who participated voluntarily.

Method

Participants were asked to rate the prominence of the adjective and the noun on a seven point scale. The task was web-based and designed using WWStim (Veenker, 2003). Stimuli consisted of html-pages on which the utterance could be played as many times as needed using a button. Utterances were presented in a random order which was different for each participant.

Prominence scores are again computed as a relative measure, for which the prominence value of the noun is subtracted from the prominence value of the adjective. This measure accounts for the fact that the perception of prominence is dependent on surrounding material in a phrase (Gussenhoven, Repp, Rietveld, Rump, & Terken, 1997). Furthermore, possible individual differences in the use of the rating scale (i.e. tendencies to use only one end of the scale) are abstracted over by such a measure. ANOVAs are performed on repeated relative prominence measures of size and color adjectives as dependent variable with communicative setting (2 levels: shared, conceal) and contrast (2 levels: single, double) as within-subject factors.

Results

Overall, the results for size adjectives are not significantly different from those for color (M_{size} = .32, M_{color} = .34, $t(2923) = -.24$, p = .81). Results indicate that adjectives in the conceal setting were perceived with more prominence (M_{size} = .90, M_{color} = .78) than in the shared setting (M_{size} = -.36, M_{color} = -.01), see Table 4 and Figure 5. Pearson's correlation coefficient indicates that the relative prominence scores closely resemble the relative pitch measures: [size: r = .45, n = 101, $p < .001$; color: r = .67, n = 120, $p < .001$].

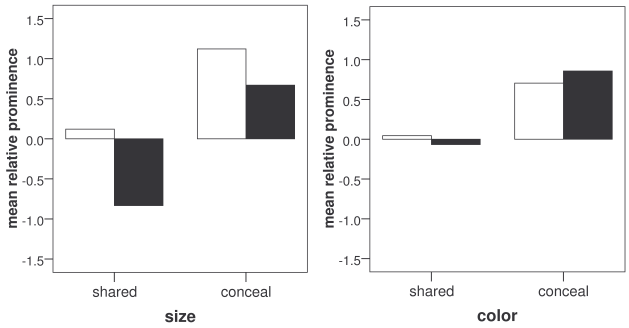

Figure 5: Mean relative prominence for size (left) and color (right) adjectives as a function of communicative setting and contrast (white = single, black = double).

Conclusion

This study shows that the more salient a contrast is between figures (i.e. by communicative setting, contrast or feature) the more likely speakers are to refer to a contrast with an adjective. There is a clear division between matching and mismatching adjectives in that only the former are affected by salience factors. Such a finding is in accordance with incremental speech production strategies; an adjective may be uttered when the context to which it refers is not be fully known to the speaker (i.e. mismatching). For the same reason, the salience of the target figure, which is determined by its context, affected the likelihood of usage only for contrast-matching adjectives. An exception is the general effect of feature, which is found in both contrast-matching and contrast-mismatching data.

Looking at the salience factors one by one, we can make a division between factors that are clear from the visual representation of a contrast only (type of contrast and feature) and factors for which additional cognitive processing is needed (communicative setting). Only the latter type affects the lexical *and* the prosodic behaviour of the speaker in that the adjective is both more likely to be used and more prominent in the conceal setting.

Table 4: Effects of communicative setting and contrast on repeated measures of relative prominence on size and color adjectives after analysis of variance.

Factor	Size adjectives	Color adjectives
Setting	$F(1,12) = 20.17$, $p < .001$, $\eta_p^2 = .63$	$F(1,12) = 17.38$, $p < .001$, $\eta_p^2 = .59$
Contrast	$F(1,12) = 13.75$, $p < .01$, $\eta_p^2 = .53$	$F(1,12) < 1$, n.s.

As for salience, Brown (1983) investigated when speakers phonologically reduce given information as opposed to new information. Although it was found that speakers sometimes introduced inferable information without reduction, they always attenuate words referring to information introduced in a previous discourse context or information evocable from the physical situation (i.e. highly salient information). The current results are compatible with Brown (1983) in that prosodic behaviour is affected by what is cognitively salient in the mind of the speaker and not necessarily by what is visually salient in a certain context.

The correlation between F0 measures and prominence ratings suggest that the acoustic cue listeners relied upon was pitch. Note that other work found that loudness, which is not taken into account here, plays the major role in the perception of prominence (Kochanski, Grabe, Coleman, & Rosner, 2005). Pitch is however, unlike in word stress, key in the realization of pitch accents. Interestingly, in the current experimental setup the contrast relation between two figures spanned one visual discourse context. This is the type of context for which Pechmann (1984) finds that it is unlikely that speakers mark a contrast by means of accentuation, which he explains in terms of incremental speech production.

Nevertheless, the present prosodic data are compatible with incremental production strategies (Fry, 1969; Levelt, 1989; Pechmann, 1989). That is, the prohibitive instruction may oblige speakers to pay attention to the contrastive occluded figure before they start articulating. In other words, speakers plausibly have a cognitive representation of the contrast relation upon articulation in the conceal setting and not in the shared setting. Such an assumption could be supported by data from eye-tracking or speech onset times, which are left for future research.

To conclude, Wegner and colleagues (1987, 1994) show that when people are instructed not to think of a pink elephant, they cannot avoid doing so. Wardlow Lane et al. (2006) find the same effect in speakers' use of adjectives when instructed to ignore certain information. The present study extends this finding by showing that the more salient a feature of a picture, the more often an adjective is used to refer to it. It does not matter whether the feature derives its salience from inherent characteristics (size or color), its relation to other figures (single or double contrast) or whether speakers' attention is drawn towards it (shared, conceal). However, only salience which is related to speakers' attention affected also the adjective's *prosodic* realisation. That is, when a feature becomes cognitively rather than visually more salient, its reference by means of an adjective is acoustically more prominent.

Acknowledgements

We thank Ruud Koolen, Jorrig Vogels and four anonymous reviewers for valuable comments on previous versions of this paper as well as all participants for their effort.

References

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech, 47*, 31-56.

Boersma, P., & Weenink, D. (2010). Praat: doing phonetics by computer (Version 5.1.25).

Brown, G. (1983). Prosodic structure and the given / new distinction. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and Measurements* (pp. 67–78). New York: Springer-Verlag.

Clark, H. H. (1996). *Using language*. Cambridge; New York: Cambridge University Press.

Collier, R., & Hart, J. t. (1981). *Cursus Nederlandse intonatie*. Leuven: Acco.

Fry, D. B. (1969). The linguistic evidence of speech errors. *BRNO Studies of English, 8*, 69-74.

Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research, 47*, 103-138.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics*. New York: Academic Press.

Gussenhoven, C., Repp, B. H., Rietveld, A., Rump, H. H., & Terken, J. (1997). The perceptual prominence of fundamental frequency peaks. *Journal of the Acoustical Society of America*, *102*(5), 3009–3022.

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: fundamental frequency lends little. *Journal of the Acoustical Society of America, 118*(2), 1038-1054.

Levelt, W. J. M. (1989). Speaking : from *intention to articulation*. Cambridge, Mass.: MIT Press.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech, 6*, 172–187.

Pechmann, T. (1984). *Überspezifizierung und Betonung in referentieller Kommunikation.* Unpublished Dissertation, Universität Mannheim, Mannheim.

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics, 27*, 89-110.

Veenker, T. J. G. (2003). WWStim: A CGI script for presenting web-based questionnaires and experiments (Version 1.4.4): Utrecht University.

Wardlow Lane, L., Groisman, M., & Ferreira, V. S. (2006). Don't talk about pink elephants!: speakers' control over leaking private information during language production. *Psychological Science, 17*(4), 273-277.

Wegner, D. M., Schneider, D. J., Carter, S., & White, T. (1987). Paradoxical effects of thought suppression. *Journal of Personality and Social Psychology, 53*, 5-13.

Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review, 101*, 34-52.